A Statistical Analysis of Mathematics Placement Scores

By Carlos Cantos, Anthony Rhodes and Huy Tran, under the supervision of Austina Fong

Portland State University, Spring 2014

Summary & Objectives

The general objective of the present study is to determine the efficacy of the undergraduate mathematics placement examination process at Portland State University. The data consists of 1,490 total undergraduate-level placement examination 'profiles', recorded between spring of 2013 and winter of 2014. In each case, students completed the adaptive ALEKS placement test comprising 25 to 30 questions. At the conclusion of the test, students receive a composite assessment including a *total score* in addition to a profile of eleven scores in specific sub-categories such as: trigonometry, relations and functions, exponentials and logarithms, *et al.* In the current incarnation, the ALEKS placement examination is not proctored and students are permitted to take the test remotely; the exam may be retaken without penalty an indefinite number of times. The data profiles for this study included total and sub-category scores, the course into which the student was placed, the grade of the student in said course, the overall gpa of the student, number of re-takes of the ALEKS test (if relevant), and the instructor of the placement course. Based upon their overall ALEKS score, students are placed as follows:

Score	Course Placement
75%-100%	MTH 251: Calculus I MTH 261: Linear Algebra
60%-74%	MTH 112: Introductory College Mathematics II
45%-59%	MTH 105: Excursions in Mathematics MTH 111: Introductory College Mathematics I STAT 105: Elementary Data Analysis STAT 243: Introduction to Probability and Statistics I
30%-44%	Math 95: Intermediate Algebra
15%-29%	Math 70: Elementary Algebra
0%-14%	None

The first approach we adopted was to determine the degree to which the ALEKS placement score is an indicator of future 'success' and specifically, which sub-scores demonstrated the strongest statistical linkage with future success in each placement category (or whether, conversely, any statistical correlation exists at all between ALEKS scores and future success). In fact, after exhaustive analysis, we found that *no significant* statistical correlation (using a basic Pearson correlation) exists between placement score outcomes and future success. In each case the data was first divided according to the placement-level of a particular student; in separate instances we used both a quantified grade value and a simple indicator variable (C- and above for pass) and found the correlation to be universally weak – even negative in some instances. A typical result corroborating this weak correlation is shown below for the set of all students placed into MTH 243, which yielded largest sample of all the placement courses.

Pearson Correlation Coefficients, N = 148 Prob > |r| under H0: Rho=0										
	GradeValue	Numbers	Equations	Functions	Polynomials	Rationals	Radicals	Logarithms	Trigonometry	ALEKS
GradeValue	1.00000	0.08653	0.11977	0.13428	0.18828	0.09801	0.14854	0.09915	0.13566	0.14056
		0.2957	0.1471	0.1037	0.0219	0.2360	0.0716	0.2305	0.1002	0.0884
Numbers	0.08653	1.00000	0.80630	0.71232	0.76167	0.62978	0.68259	0.46630	0.63438	0.73856
Numbers	0.2957		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Equations	0.11977	0.80630	1.00000	0.92650	0.94706	0.88018	0.88629	0.74159	0.86802	0.95769
Equations	0.1471	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Functions	0.13428	0.71232	0.92650	1.00000	0.90863	0.88971	0.86609	0.79448	0.90263	0.96534
Functions	0.1037	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Polynomials	0.18828	0.76167	0.94706	0.90863	1.00000	0.87421	0.88092	0.71218	0.84289	0.94301
Polynomials	0.0219	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001
Rationals	0.09801	0.62978	0.88018	0.88971	0.87421	1.00000	0.90865	0.85420	0.90227	0.95232
Rationals	0.2360	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001
Radicals	0.14854	0.68259	0.88629	0.86609	0.88092	0.90865	1.00000	0.81271	0.86860	0.93979
Radicals	0.0716	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001
Logarithms	0.09915	0.46630	0.74159	0.79448	0.71218	0.85420	0.81271	1.00000	0.85665	0.85899
Logarithms	0.2305	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001
Trigonometry	0.13566	0.63438	0.86802	0.90263	0.84289	0.90227	0.86860	0.85665	1.00000	0.95009
Trigonometry	0.1002	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001
ALEKS	0.14056	0.73856	0.95769	0.96534	0.94301	0.95232	0.93979	0.85899	0.95009	1.00000
ALEKS	0.0884	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	

In light of these somewhat surprising findings, we attempted to further eliminate sources of (unwanted) variance by conditioning student placement scores based upon a common instructor (in order to remove possible grading disparities); furthermore, we removed instances of placement scores generated by repeated attempts from each placement subset (in general the first placement score is most indicative of a student's 'true' subject-acumen). Once again, statistical correlation between ALEKS scores and the eventual grade of the student for the course in which they were placed was weak to non-existent. These results are summarized just below.

Mth70: 15-29

Pearson Correlation Coefficients, N = 75 Prob > |r| under H0: Rho=0														
	GradeValue	Numbers	Geometry	Inequalities	Equations	Functions	Exponents	Polynomials	Rationals	Radicals	Logarithms	Trigonometry	ALEKS	GPA
GradeValue	1.00000	0.14490	0.17235	0.24352	-0.03654	-0.14943	0.14228	.	-0.04207	0.29839	.	0.10253	0.22542	0.66242
		0.2148	0.1392	0.0353	0.7556	0.2007	0.2233	.	0.7201	0.0093	.	0.3814	0.0518	<.0001

Mth95: 30-44

Pearson Correlation Coefficients, N = 128
Prob > |r| under H0: Rho=0

	GradeValue	Numbers	Geometry	Inequalities	Equations	Functions	Exponents	Polynomials	Rationals	Radicals	Logarithms	Trigonometry	ALEKS	GPA
GradeValue	1.00000	0.06622	0.01633	-0.05186	0.03215	-0.03448	0.06876	-0.01992	0.11763	0.01730	-0.02729	-0.06109	0.02578	0.72503
		0.4577	0.8548	0.5610	0.7187	0.6992	0.4406	0.8234	0.1861	0.8463	0.7598	0.4933	0.7727	<.0001

Mth111: 45-59

Pearson Correlation Coefficients, N = 98
Prob > |r| under H0: Rho=0

	GradeValue	Numbers	Geometry	Inequalities	Equations	Functions	Exponents	Polynomials	Rationals	Radicals	Logarithms	Trigonometry	ALEKS	GPA
GradeValue	1.00000	0.08429	0.07861	0.16683	0.08832	0.21211	0.09000	0.14698	0.03384	-0.08214	0.01725	0.15475	0.12860	0.76884
		0.4093	0.4417	0.1006	0.3872	0.0360	0.3782	0.1487	0.7408	0.4214	0.8661	0.1281	0.2069	<.0001

Stat243: 45-59

Pearson Correlation Coefficients, N = 104
Prob > |r| under H0: Rho=0

	GradeValue	Numbers	Geometry	Inequalities	Equations	Functions	Exponents	Polynomials	Rationals	Radicals	Logarithms	Trigonometry	ALEKS	GPA
GradeValue	1.00000	-0.06884	0.08187	0.07551	-0.03215	0.16761	-0.01485	-0.07111	0.02824	0.00070	-0.01849	0.04067	0.03073	0.78431
		0.4875	0.4087	0.4462	0.7459	0.0890	0.8811	0.4732	0.7760	0.9944	0.8522	0.6819	0.7568	<.0001

Mth112: 60-74

Pearson Correlation Coefficients, N = 59
Prob > |r| under H0: Rho=0

	GradeValue	Numbers	Geometry	Inequalities	Equations	Functions	Exponents	Polynomials	Rationals	Radicals	Logarithms	Trigonometry	ALEKS	GPA
GradeValue	1.00000	.	-0.10447	0.11149	-0.01765	-0.02888	-0.17140	-0.08796	-0.07188	-0.07216	0.03830	0.07319	-0.05377	0.82066
		.	0.4310	0.4005	0.8945	0.8281	0.1943	0.5077	0.5885	0.5870	0.7733	0.5817	0.6858	<.0001

In an effort to potentially increase correlation, we next tried binning the data using variable bin sizes. The resultant correlation values were comparable to the non-binned data, as the reader will note.

Bin Size 5:

Pearson Correlation Coefficients, N = 74
Prob > |r| under H0: Rho=0

	GradeValue	Numbers	Geometry	Inequalities	Equations	Functions	Exponents	Polynomials	Rationals	Radicals	Logarithms	Trigonometry	ALEKS	Bin
GradeValue	1.00000	0.13216	0.16414	0.23117	-0.03662	-0.14593	0.13360	.	-0.05466	0.28242	.	0.10448	0.21108	0.18767
		0.2617	0.1623	0.0475	0.7567	0.2147	0.2565	.	0.6437	0.0148	.	0.3757	0.0710	0.1093

Bin Size 3:

Pearson Correlation Coefficients, N = 74
Prob > |r| under H0: Rho=0

	GradeValue	Numbers	Geometry	Inequalities	Equations	Functions	Exponents	Polynomials	Rationals	Radicals	Logarithms	Trigonometry	ALEKS	Bin
GradeValue	1.00000	0.13216	0.16414	0.23117	-0.03662	-0.14593	0.13360	.	-0.05466	0.28242	.	0.10448	0.21108	0.18845
		0.2617	0.1623	0.0475	0.7567	0.2147	0.2565	.	0.6437	0.0148	.	0.3757	0.0710	0.1078

Bin Size 2:

Pearson Correlation Coefficients, N = 74
Prob > |r| under H0: Rho=0

	GradeValue	Numbers	Geometry	Inequalities	Equations	Functions	Exponents	Polynomials	Rationals	Radicals	Logarithms	Trigonometry	ALEKS	Bin
GradeValue	1.00000	0.13216	0.16414	0.23117	-0.03662	-0.14593	0.13360	.	-0.05466	0.28242	.	0.10448	0.21108	0.20305
		0.2617	0.1623	0.0475	0.7567	0.2147	0.2565	.	0.6437	0.0148	.	0.3757	0.0710	0.0827

These persistent weak correlation values provide strong evidence that, under the current method, ALEKS placement exams scores are not necessarily indicative of future success in the given course recommended *via* the placement assessment. This is, however, not to suggest that the placement exam is entirely ineffectual as it is currently implemented. Instructor Fong has, for instance, demonstrated in her own research that students who place into a given course using the ALEKS test (as opposed to placing with prerequisites alone) are much more likely to succeed in that course than those who do not.

Our next approach was to determine whether changing the cut-off range used to place a student would lead to increased correlation values, a spike in future student success rates and the like. To this end, we analyzed a series of ROC (receiver operating characteristic) curves, in which students were initially grouped according to their ALEKS placement score. At each stage of the analysis, the lower-end of the cut-off range was increased so that the sample pool became more and more rarefied. For each new cut-off score, we assessed the number of 'true positive' (i.e. ALEKS properly placed a passing student), 'false positive' (ALEKS placed a non-passing student), 'true negative' and 'false negative' results. Using a 'confusion matrix' for each stage of this procedure, it is often possible to determine an ideal threshold value that maximizes the proportion of 'true' subjects as assessed – in this instance – by the ALEKS placement test. However, for the current data, each ROC curve was highly 'irregular' (i.e. exhibiting a non-uniform concavity and unsuitable shape) and so these results were also not statistically significant.

Building on the notion of altering the ALEKS placement cut-off scores in an effort to identify some semblance of statistical correlation between exam score and future success, we next considered an analysis of the pass and drop rates of placement levels as the cut-off range varies. In some instances this approach generated positive correlation (albeit still a relatively weak value, with $r \approx .20$); using a maximal increase in the average rate of change of the ratio of the pass to drop rate, the various cut-off scores were ranked accordingly.

The results that follow represent an 'ensemble' of variable ALEKS cut-off scores, ranked by maximal pass-to-fail rate ratio. In the analysis, we considered combinations of sub-scores and the extent to which these combinations influence overall pass rates. Note that the overall cut-off score remains unchanged. As before, if a student no longer satisfies the new cut-off score they are dropped from the simulated course and then, subsequently, new pass and drop rates are calculated for the simulation. The cut-off values are increased in increments of five; all combinations are simulated, with the exception of those containing zero as an individual sub-category cut-off. In the tables below, each column represents a simulated combination in which the top three rows give the cut-off score for each sub-category, followed be the resulting pass rate, drop rate and ratio. The top ten results (based upon p/d ratio) are given, so that a human reader can, by inspection, dismiss any practically undesirable results (e.g. drop rate too high). The wedges for each course were chosen according to guidance offered by instructor Fong, who used practical knowledge in conjunction with knowledge of the averages of the various sub-categories for each class.

First, the algorithm is run with no restrictions (except that a cut-off score of zero is disallowed). However, the results are, predictably, that very low cutoff scores result in low drop rates (MTH095 in particular). Consequently, those results are presented first, but following these findings we attempted to restrict cut-off scores so that they were generally closer to the empirical mean scores. These ranked test scores might be loosely considered as recommended cut-off regions for future course placement.

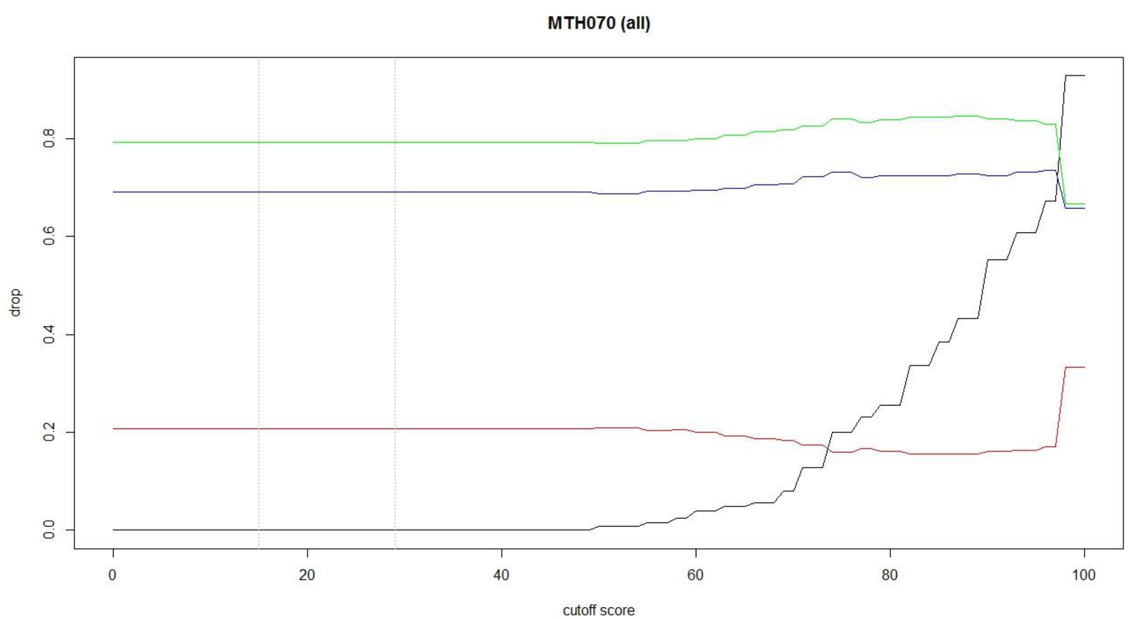
SINGLE SUBJECT SCORE CUTOFF CHARTS

Black: drop rate / Green: pass rate / Red: fail rate / Blue: mean grade (GPA score divided by 4)

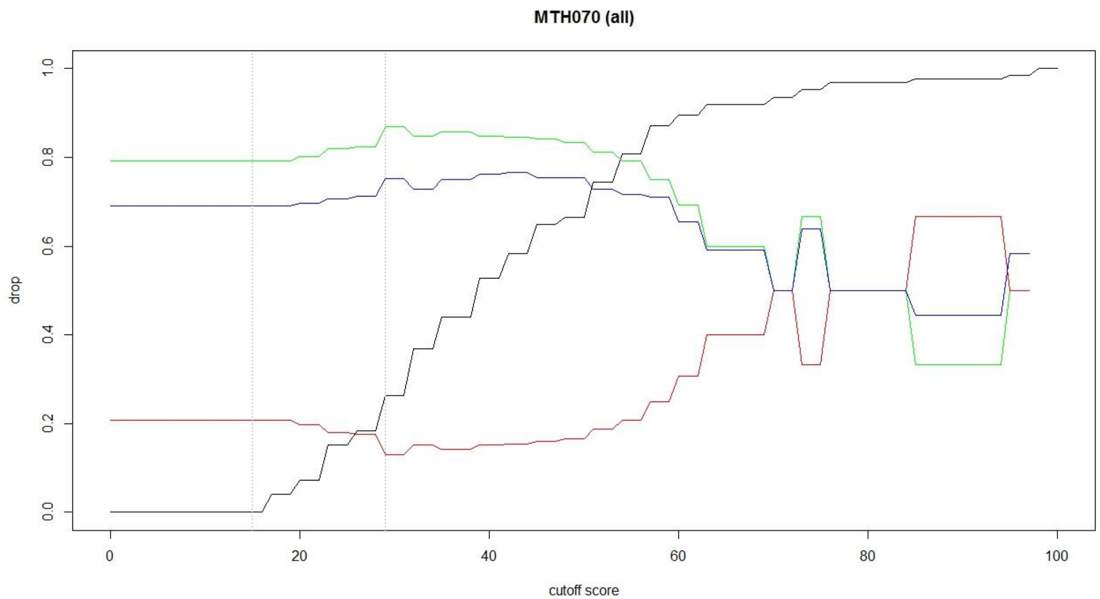
(Ignore grey lines)

Only those charts which appear to show some gain by increasing the cutoff score, without excessive loss of students, are presented.

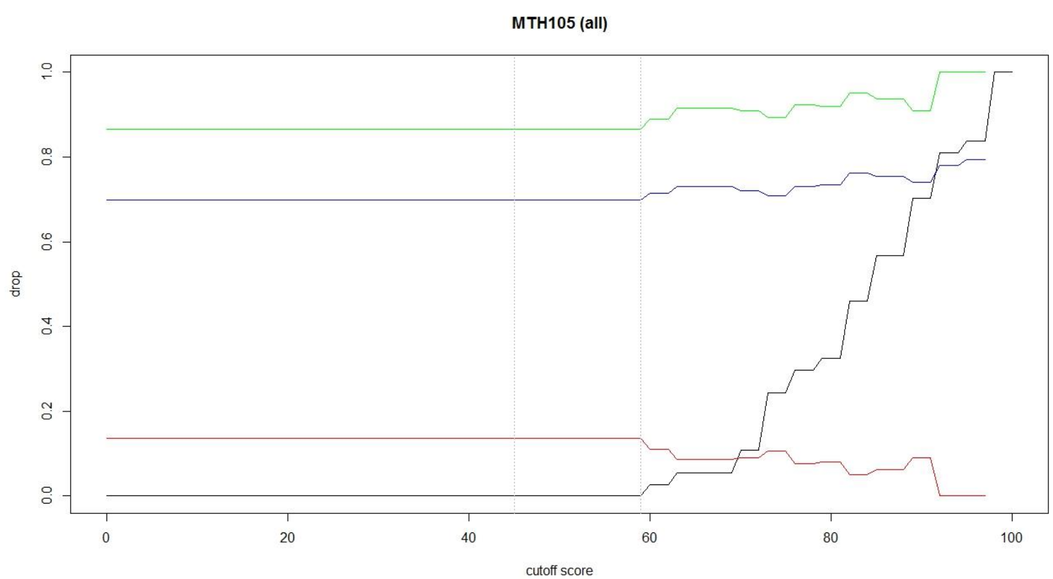
WHOLE NUMBERS, FRACTIONS, AND DECIMALS



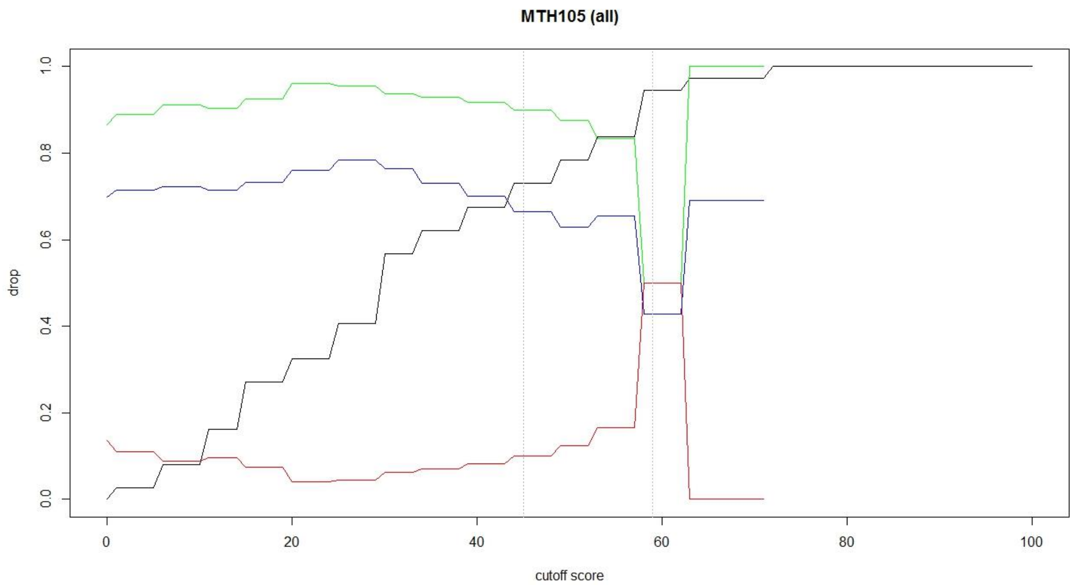
PERCENTS, PROPORTIONS, AND GEOMETRY



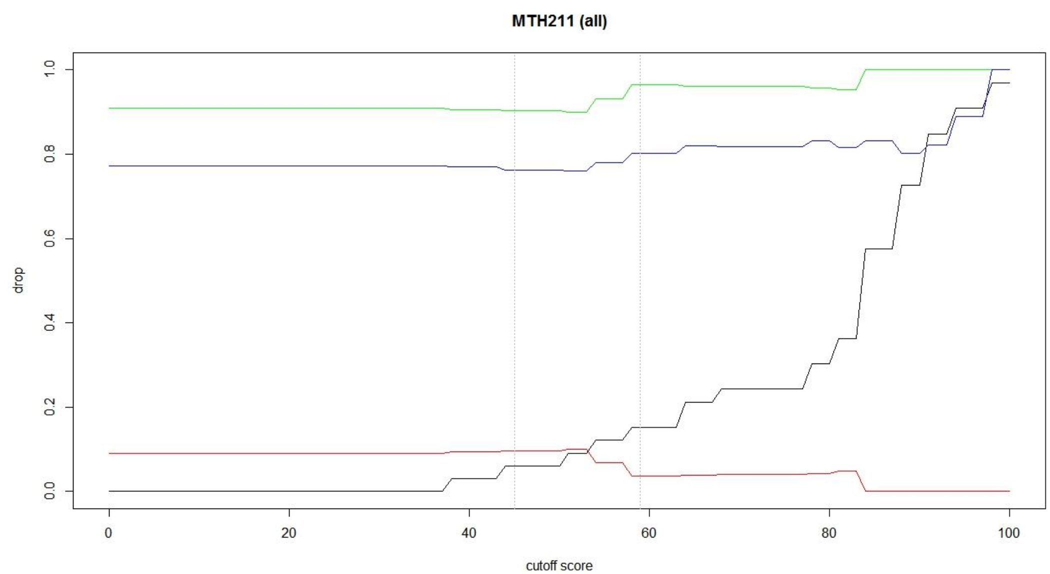
PERCENTS, PROPORTIONS, AND GEOMETRY



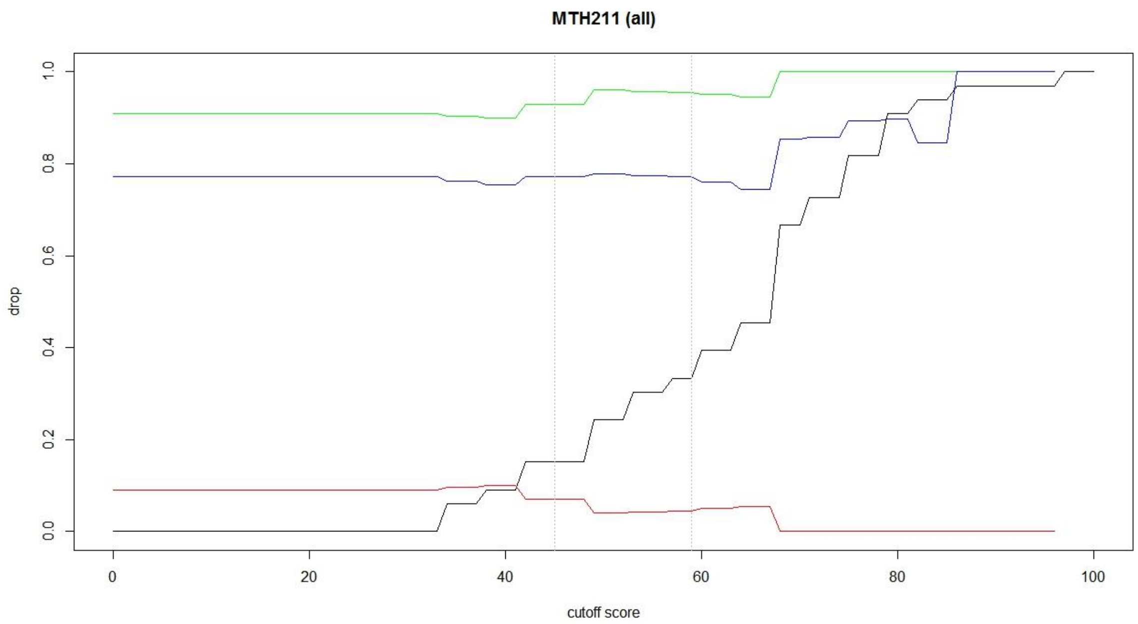
QUADRATIC AND POLYNOMIAL FUNCTIONS



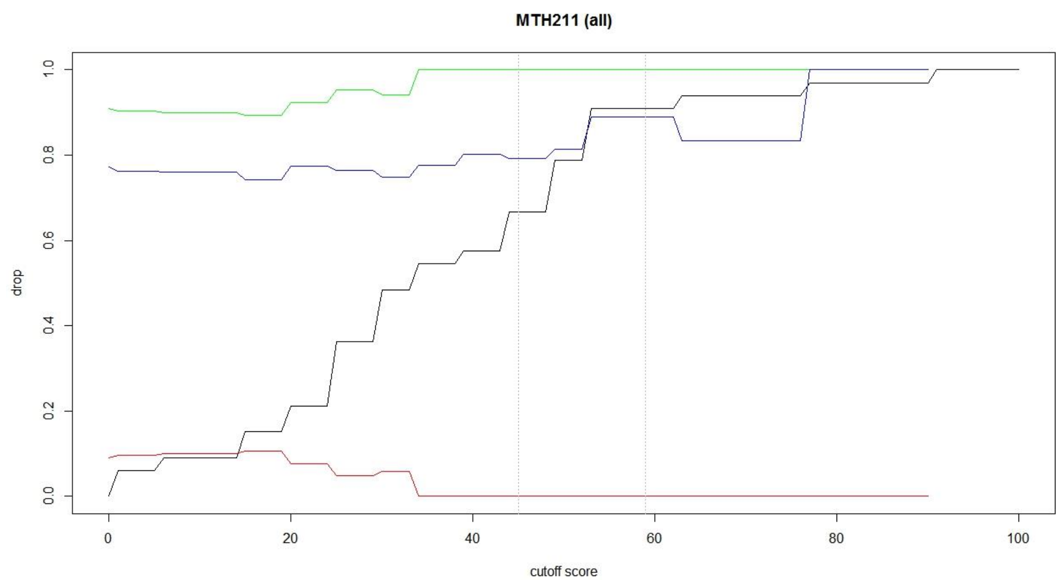
INTEGER EXPONENTS AND FACTORING



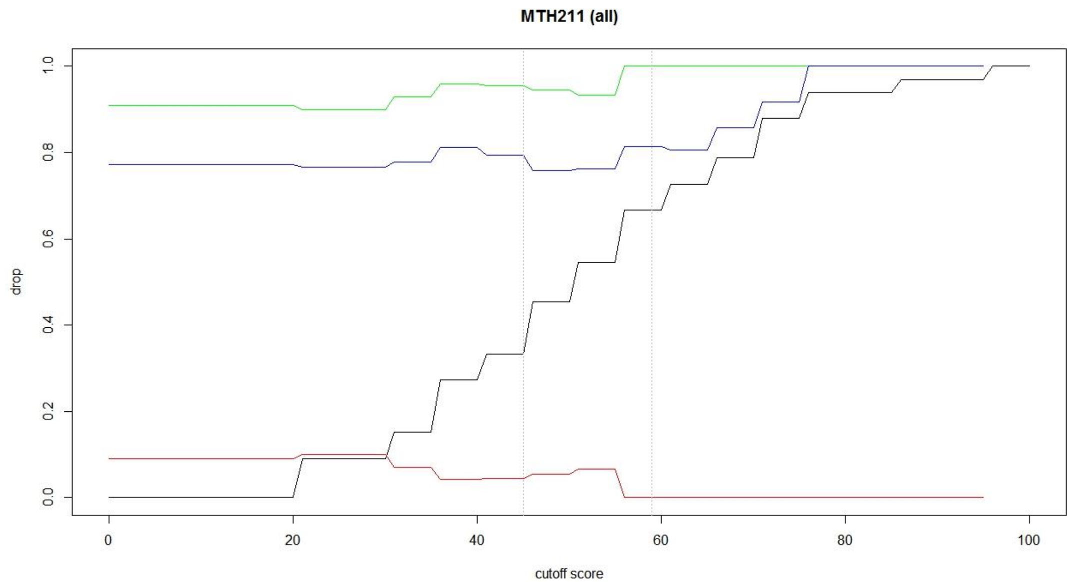
LINES AND SYSTEMS OF LINEAR EQUATIONS



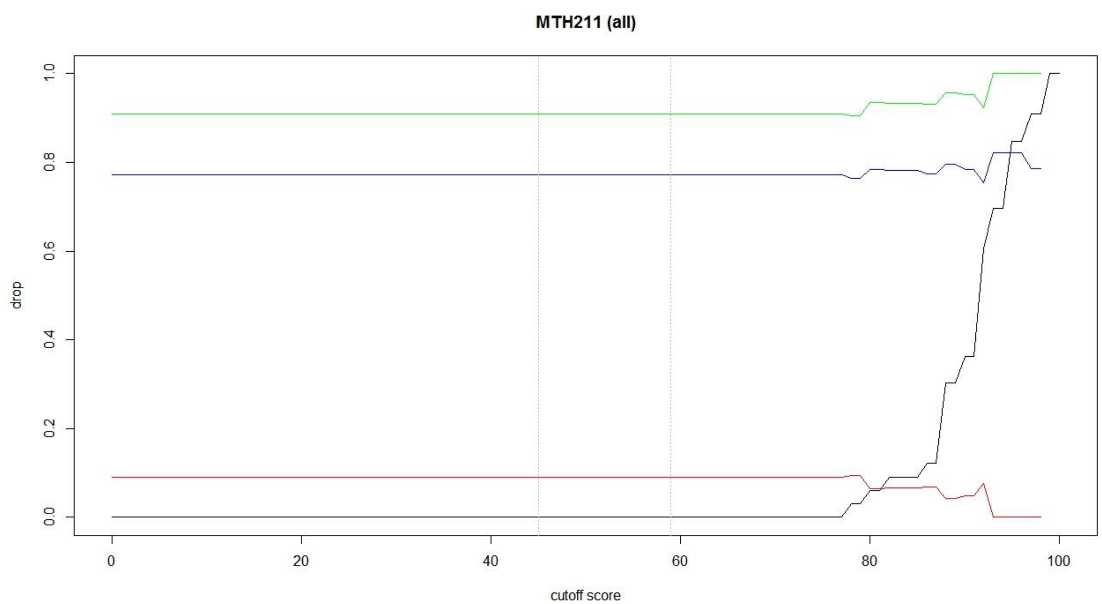
QUADRATIC AND POLYNOMIAL FUNCTIONS



RADICALS AND RATIONAL EXPONENTS



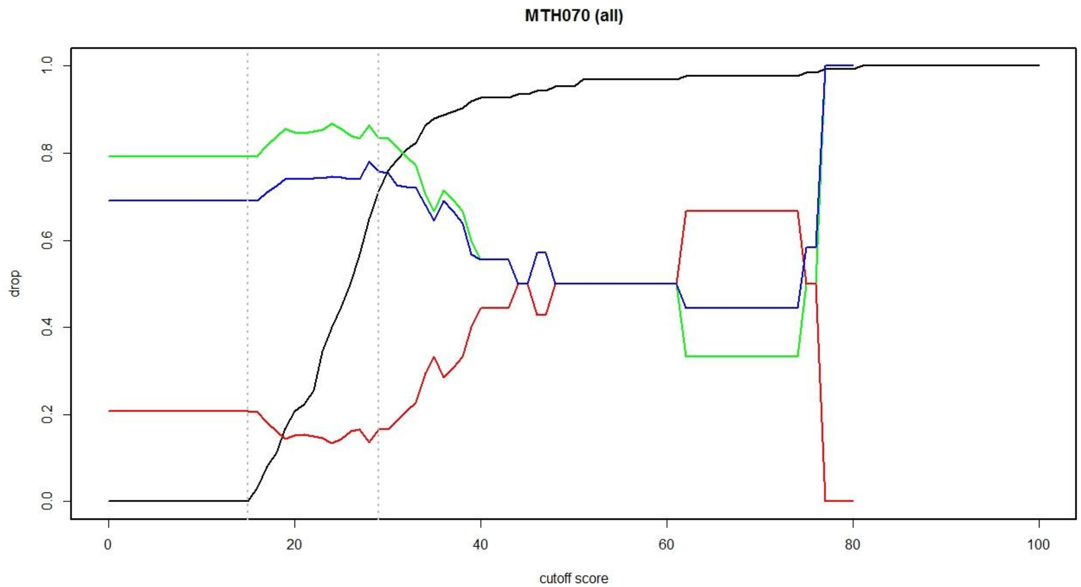
SIGNED NUMBERS, LINEAR EQUATIONS, AND INEQUALITIES

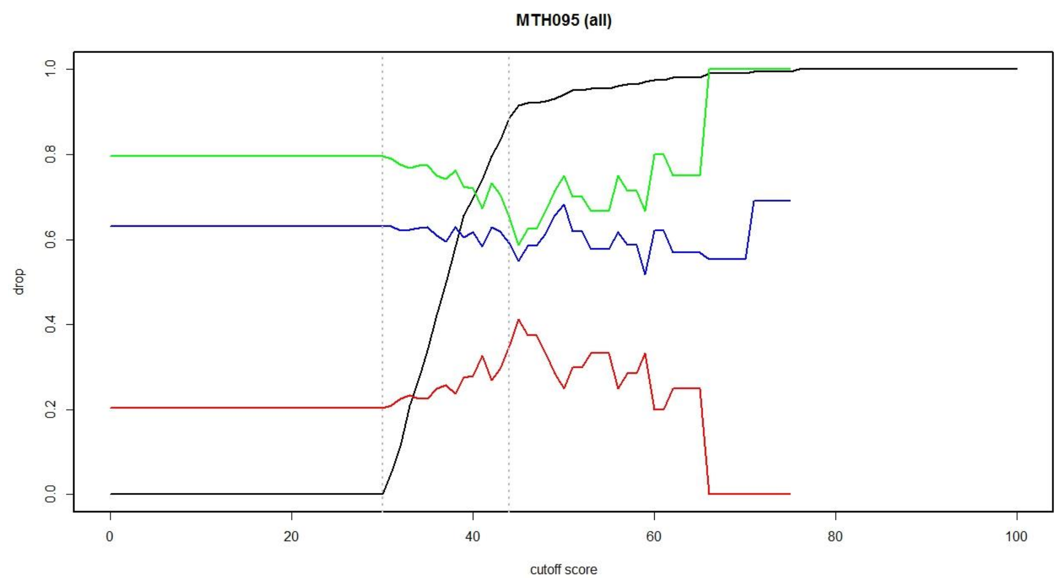


TOTAL SCORE CUTOFF CHARTS

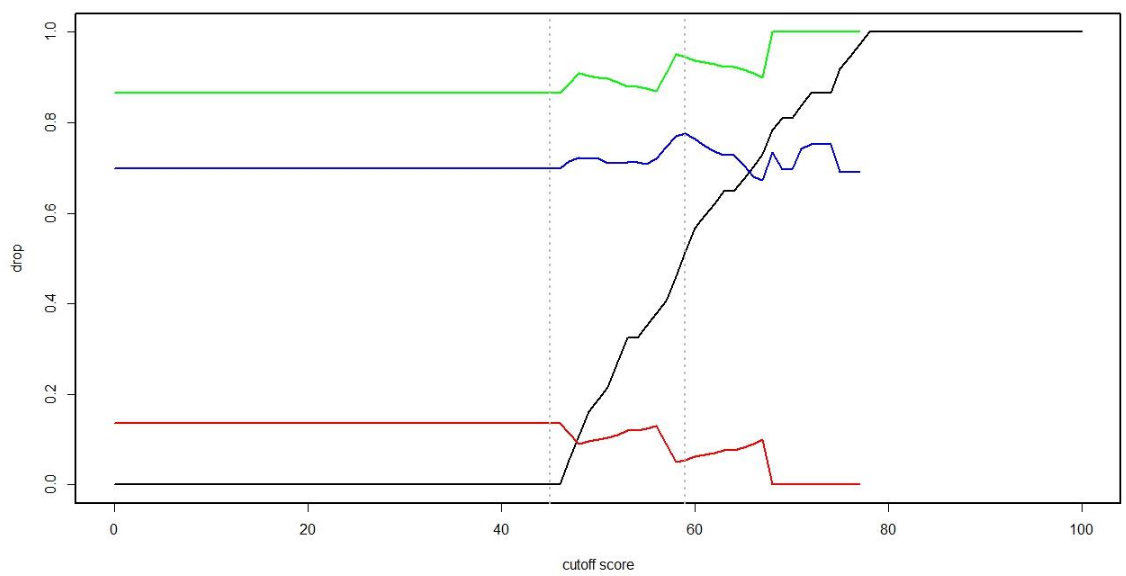
Black: drop rate / Green: pass rate / Red: fail rate / Blue: mean grade (GPA score divided by 4)

Grey: range of score placing into course

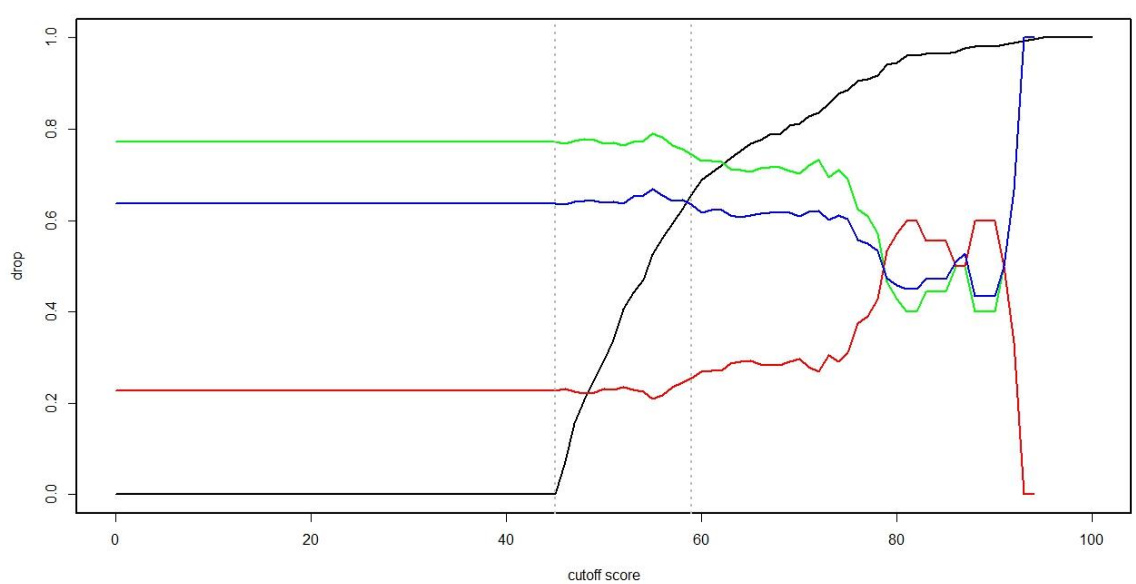


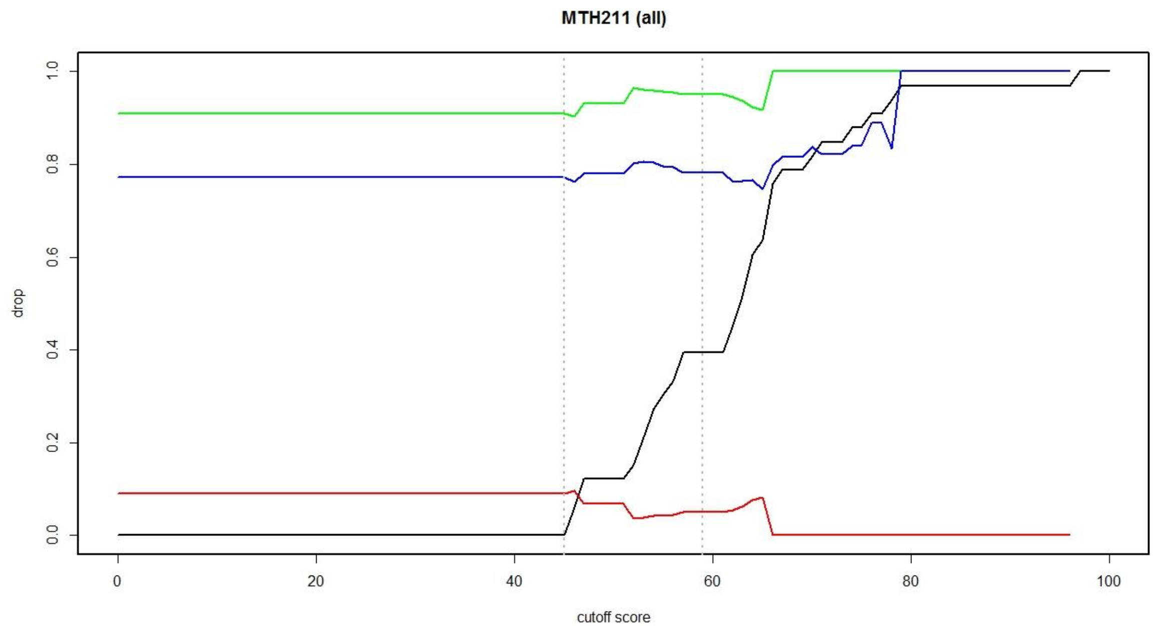


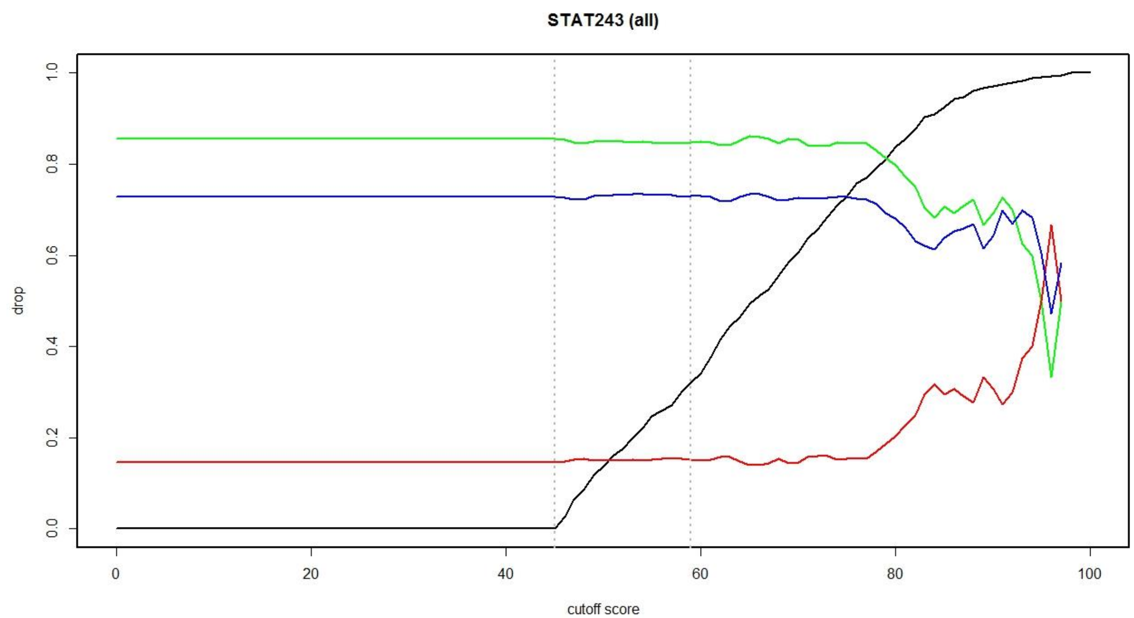
MTH105 (all)

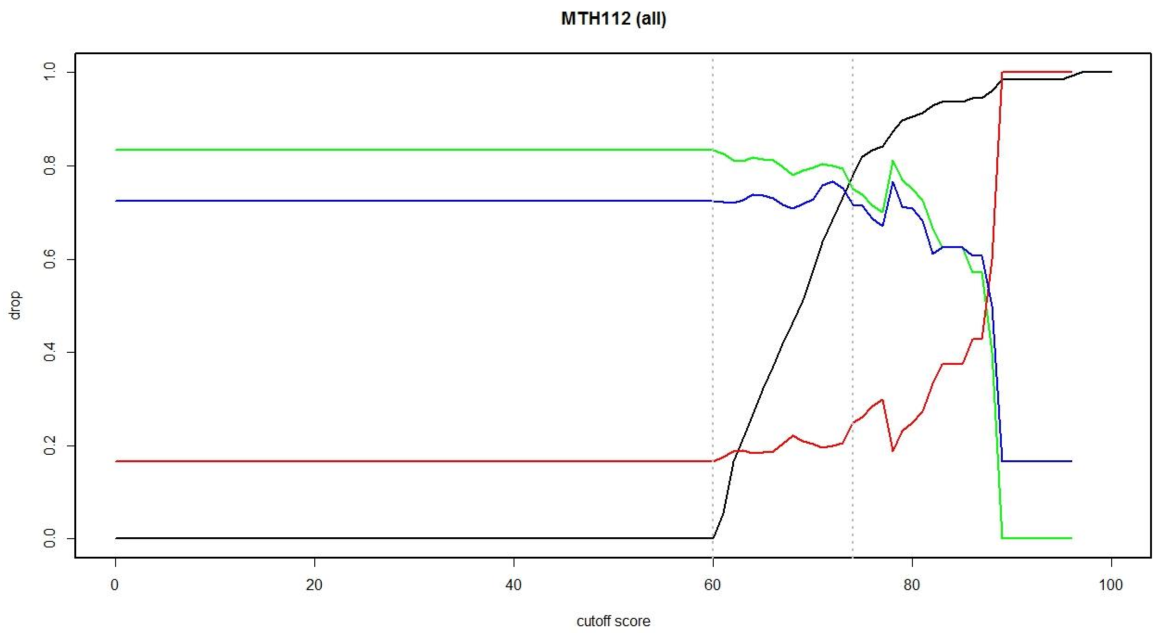


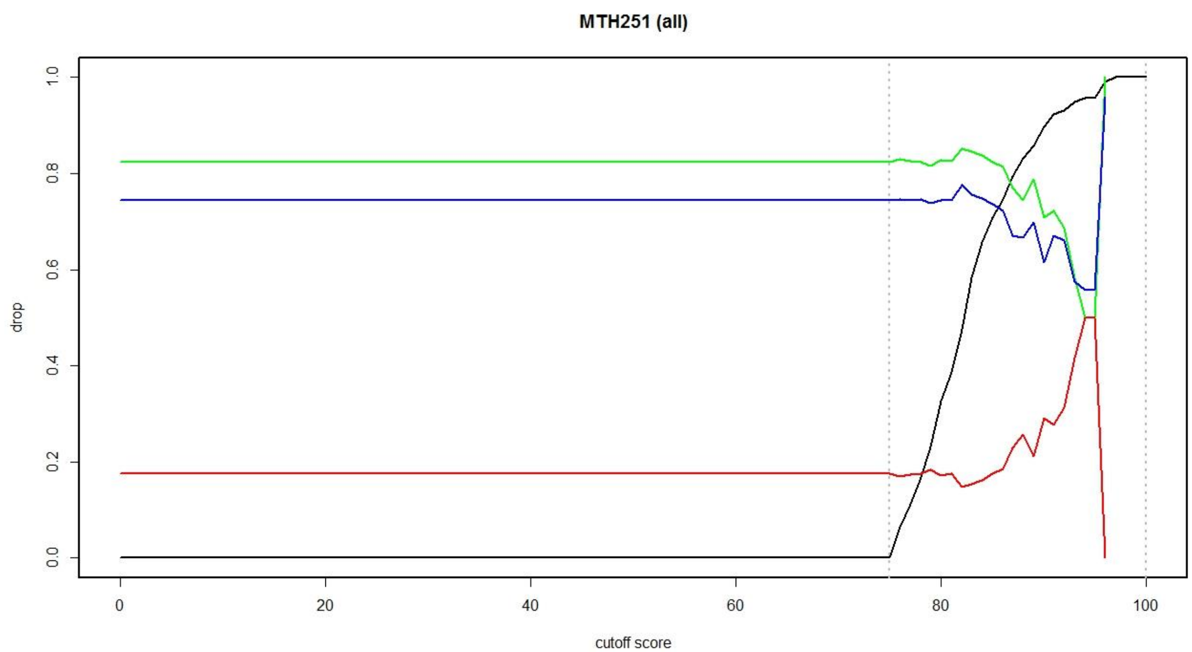
MTH111 (all)

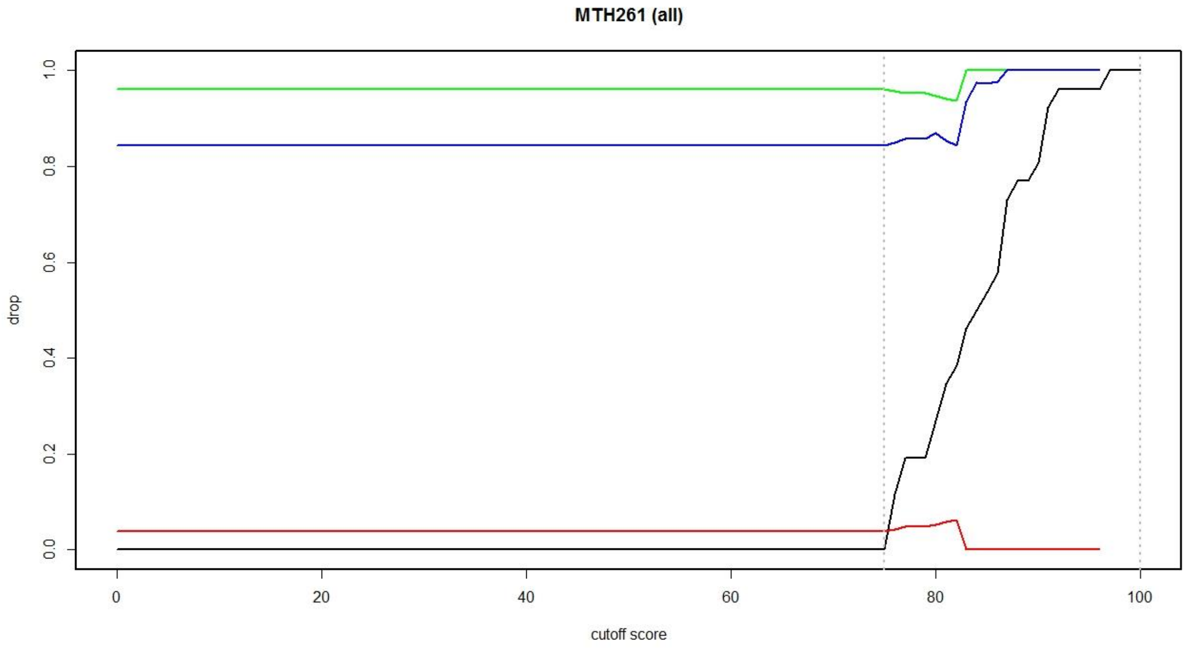



MTH211 (all)
drop
cutoff score
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100


STAT243 (all)
drop
cutoff score
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100


MTH112 (all)
drop
cutoff score
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100


MTH251 (all)
drop
cutoff score
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100



Now we look at combinations of wedges, and for each student create an indicator variable for whether they exceed each cutoff. These are summed, so that there is a vector of length n (the total unchanged class size) where each student has a number between 0 and the number of wedges (usually 3). We then calculate the correlation between this vector and the vector indicating whether they passed the course.

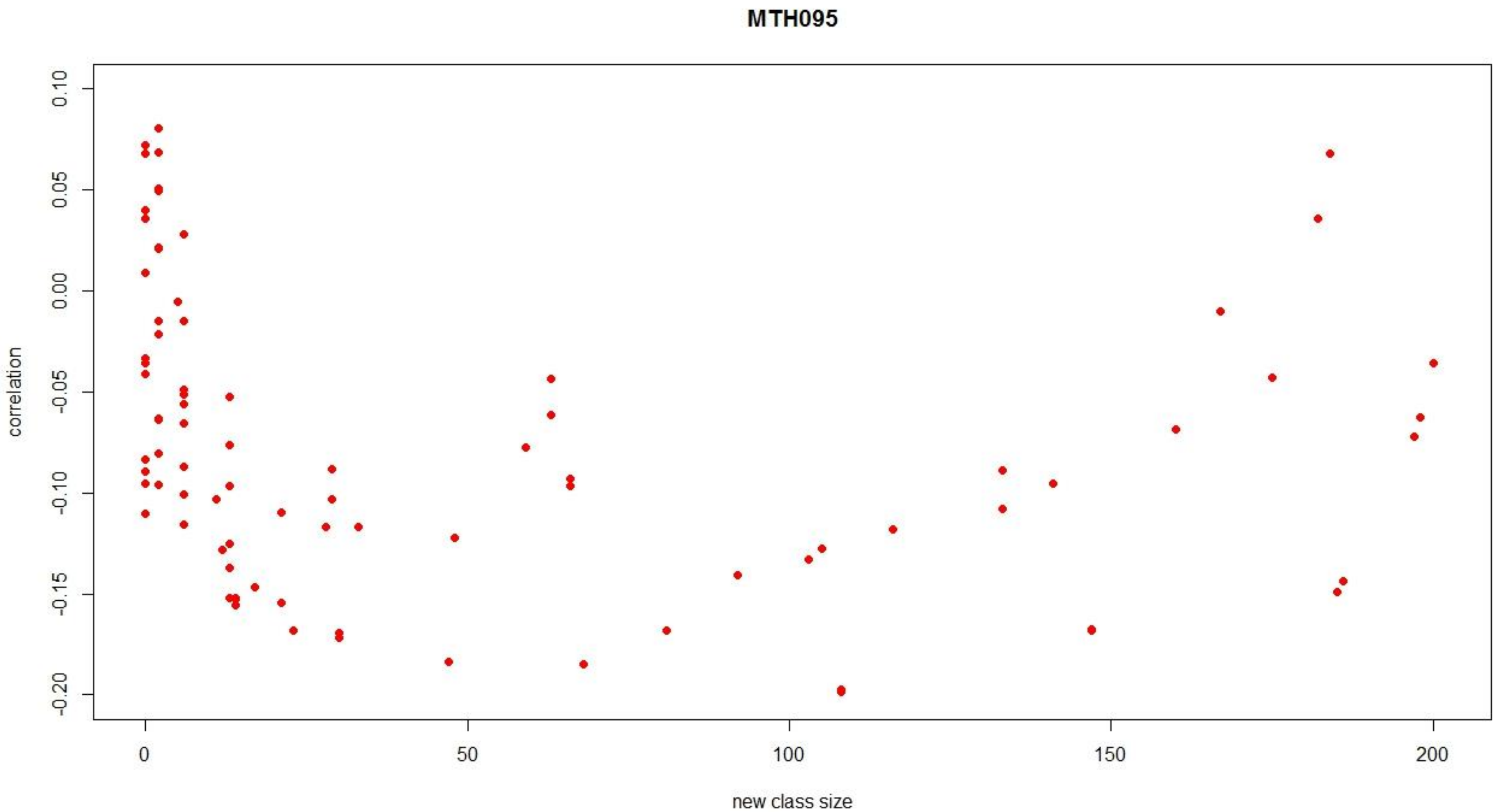
Hopefully the correlation is high for some combination, which would give some evidence towards the idea that exceeding that combination of wedge cutoffs leads to a higher chance of passing the course. We also keep track of the size of the class, so that results with very low numbers of remaining students can be ignored.

The top 10 results, ordered by correlation, are given, as well as a scatter plot of correlation versus class size. (The plot is to help scan for a good result which may not present in the top 10, since the top 10 could be full of entries with ridiculously low class sizes.)

One thing we can also see from these plots, is that as we go further in the course sequence (i.e. from 095 to 251), the density of points corresponding to low class size is much higher than those to high class size. (What is referred to is strictly the density of points, regardless of the correlation which is the main focus up to now.) This basically means that implementing any kind of non-trivial cutoff scores for these wedges causes most of the students to be dropped. This seems to be a pretty clear condemnation of the idea that ALEKS tests reliably for prerequisite knowledge, e.g. trigonometry for 251.

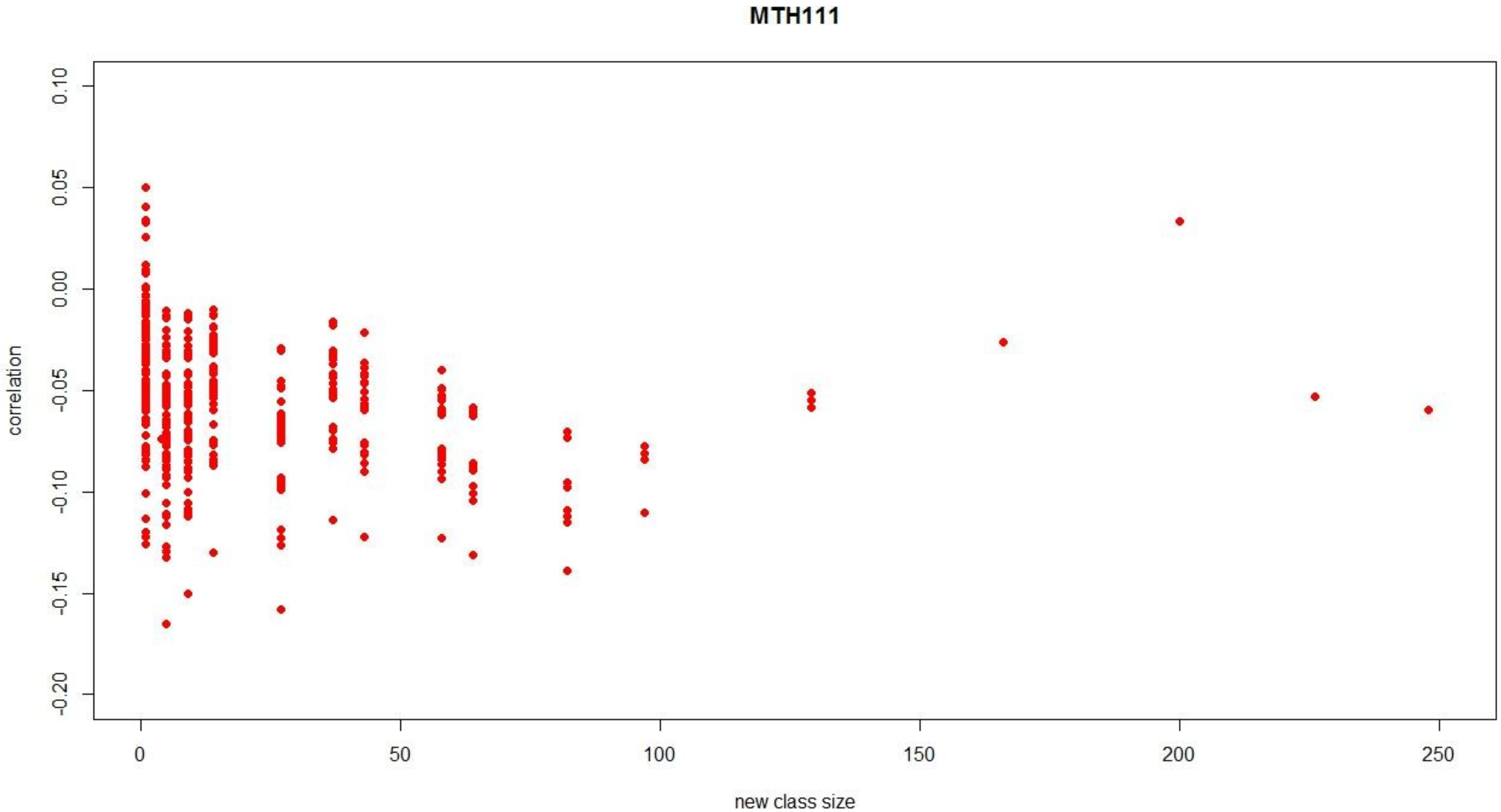
MTH095 (original pass rate = 0.79602, original class size = 201)
Wedge 1: Percents, proportions, and geometry (avg=58.3)
Wedge 2: Signed numbers, linear equations and inequalities (avg=70.4)
Wedge 3: -

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
wedge1	45.00000	90.000000	90.000000	45.000000	45.000000	5.000000	10.000000	15.000000	20.000000	25.000000
wedge2	95.00000	100.000000	95.000000	50.000000	100.000000	95.000000	95.000000	95.000000	95.000000	95.000000
pass	1.00000	NaN	1.000000	0.804348	NaN	1.000000	1.000000	1.000000	1.000000	1.000000
drop	0.99005	1.000000	0.990050	0.084577	1.000000	0.990050	0.990050	0.990050	0.990050	0.990050
size	2.00000	0.000000	2.000000	184.000000	0.000000	2.000000	2.000000	2.000000	2.000000	2.000000
corr	0.08032	0.072132	0.068361	0.067993	0.067993	0.050748	0.050748	0.050748	0.050748	0.050748



MTH111 (original pass rate = 0.77291, original class size = 251)
Wedge 1: Lines and systems of linear equations (avg=50.1)
Wedge 2: Integer exponents and factoring (avg=64.1)
Wedge 3: Radicals and rational exponents (avg=34.6)

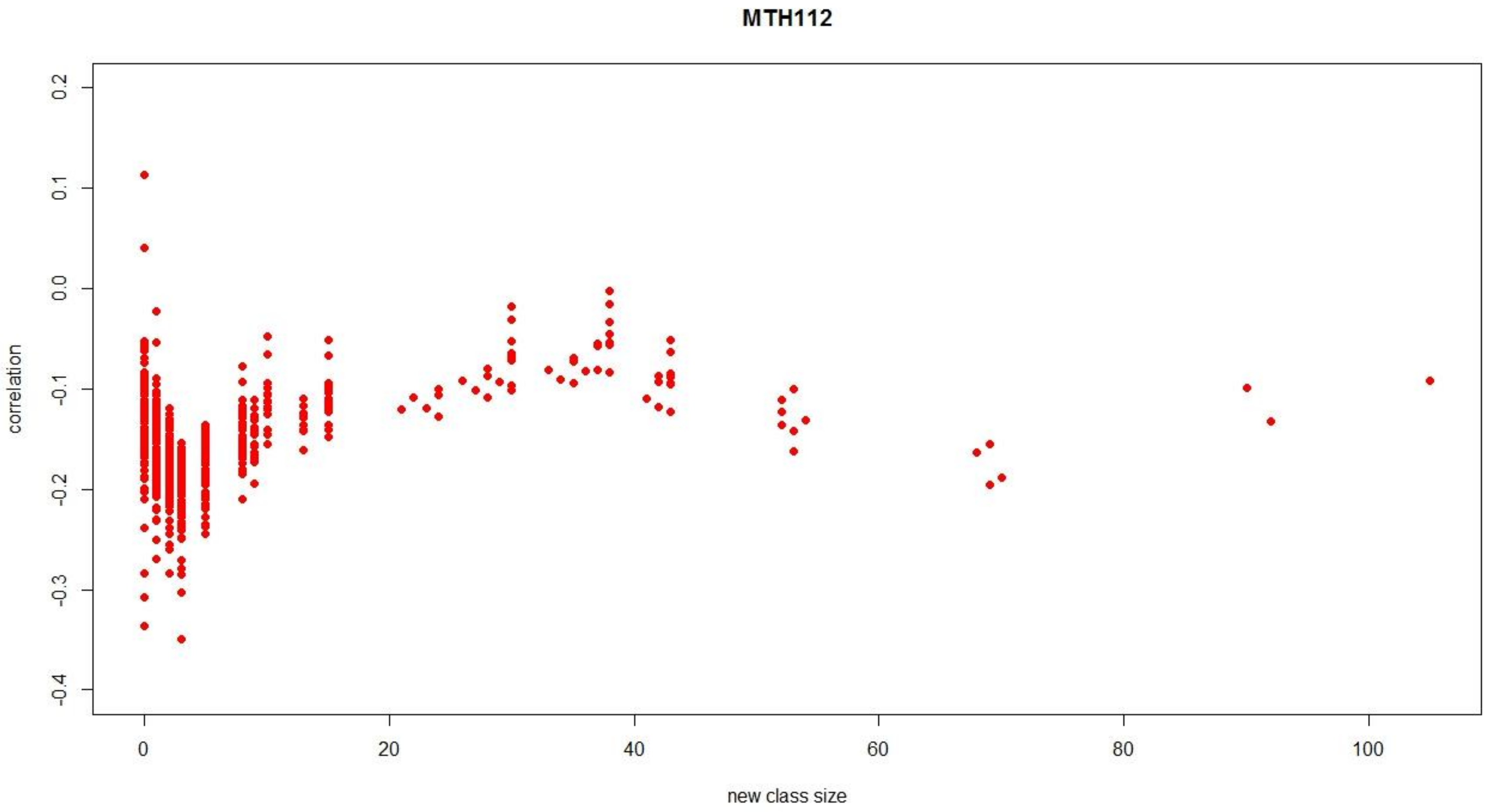
	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
wedge1	5.000000	5.000000	10.000000	10.000000	15.000000	15.000000	20.000000	20.000000	25.000000	25.000000
wedge2	45.000000	45.000000	45.000000	45.000000	45.000000	45.000000	45.000000	45.000000	45.000000	45.000000
wedge3	95.000000	100.000000	95.000000	100.000000	95.000000	100.000000	95.000000	100.000000	95.000000	100.000000
pass	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
drop	0.996016	0.996016	0.996016	0.996016	0.996016	0.996016	0.996016	0.996016	0.996016	0.996016
size	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
corr	0.049963	0.049963	0.049963	0.049963	0.049963	0.049963	0.049963	0.049963	0.049963	0.049963



As one can see, there are some combinations with larger class size and similar correlation; but, that correlation is still very low.

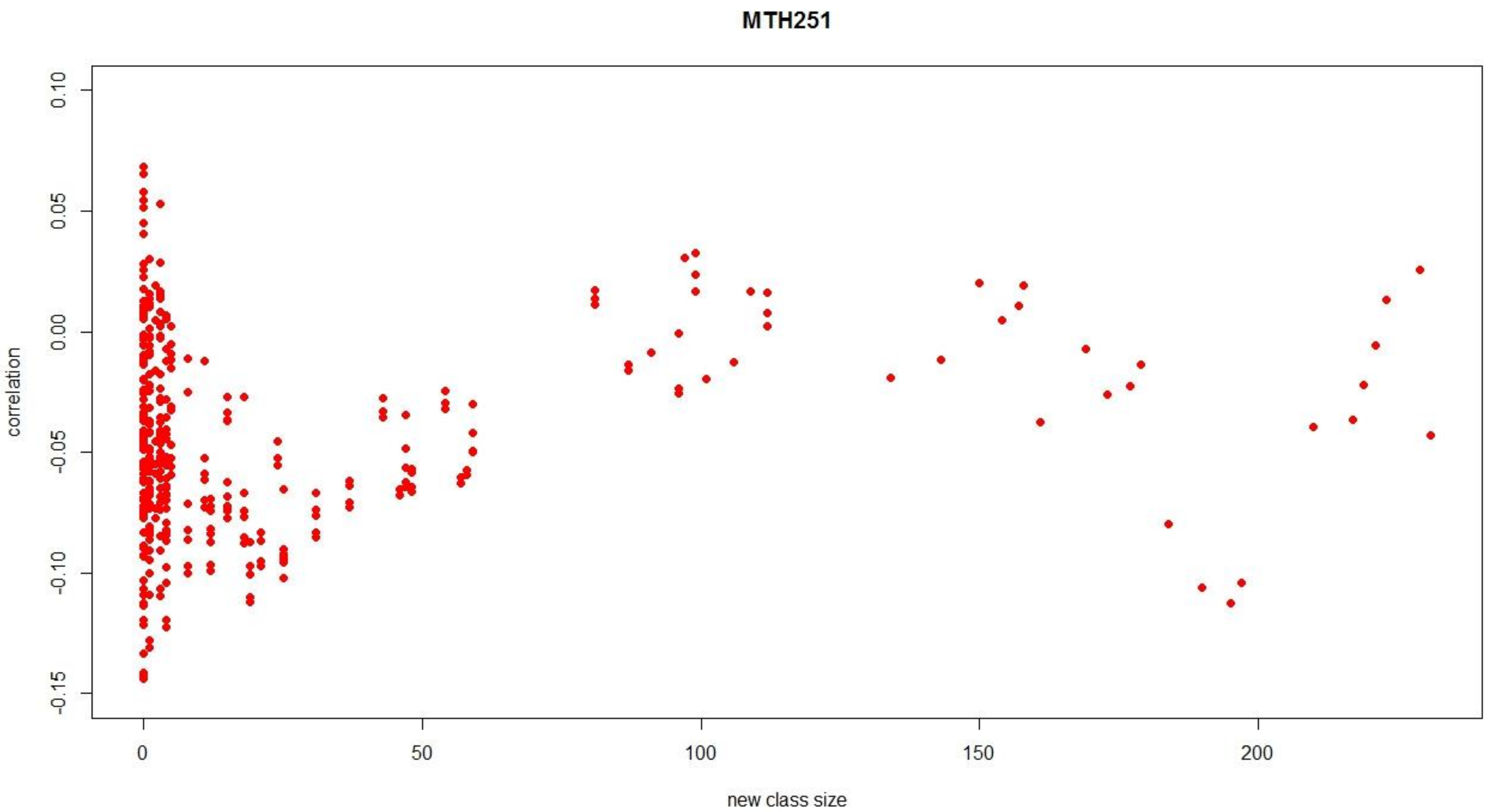
MTH112 (original pass rate = 0.83465, original class size = 127)
Wedge 1: Relations and functions (avg=25.9)
Wedge 2: Rational expressions and functions (avg=35.2)
Wedge 3: Exponentials and logarithms (avg=17.3)

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
wedge1	10.00000	10.00000	10.00000	10.000000	10.000000	10.000000	15.000000	15.000000	15.000000	10.0000000
wedge2	15.00000	15.00000	15.00000	20.000000	20.000000	20.000000	20.000000	20.000000	20.000000	15.0000000
wedge3	90.00000	95.00000	100.00000	90.000000	95.000000	100.000000	90.000000	95.000000	100.000000	40.0000000
pass	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.8157895
drop	1.00000	1.00000	1.00000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.7007874
size	0.00000	0.00000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	38.0000000
corr	0.11394	0.11394	0.11394	0.041089	0.041089	0.041089	0.041089	0.041089	0.041089	-0.0020673



MTH251 (original pass rate = 0.82403, original class size = 233)
Wedge 1: Relations and functions (avg=56.7)
Wedge 2: Exponentials and logarithms (avg=40.1)
Wedge 3: Trigonometry (avg=38.2)

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
wedge1	65.00000	65.000000	65.000000	65.000000	60.000000	5.000000	5.000000	10.000000	60.000000	60.000000
wedge2	85.00000	85.000000	85.000000	95.000000	85.000000	10.000000	15.000000	15.000000	85.000000	85.000000
wedge3	90.00000	95.000000	100.000000	100.000000	90.000000	85.000000	85.000000	85.000000	95.000000	100.000000
pass	NaN	NaN	NaN	NaN	NaN	1.000000	1.000000	1.000000	NaN	NaN
drop	1.00000	1.000000	1.000000	1.000000	1.000000	0.987124	0.987124	0.987124	1.000000	1.000000
size	0.00000	0.000000	0.000000	0.000000	0.000000	3.000000	3.000000	3.000000	0.000000	0.000000
corr	0.06814	0.065464	0.065464	0.057805	0.054596	0.052776	0.052776	0.052776	0.051501	0.051501



Again, there are some combinations with larger class size and similar correlation; but, that correlation is still very low.

Concluding Comments & Future Analysis

The general absence of a statistically significant correlation between ALEKS scores and future student success in the present data obviates our abitility to build meaningful predictive models – and building useful predictive models was our primary objective at the outset of this study. For future analysis, it is recommended that the ALEKS scores for Portland State University students under the current testing conditions are compared and contrasted with that of other, comparable groups of students at other colleges and univerities. In particular, we would like to know whether schools with more stringent testing conditions for placement exams (e.g. proctored tests, limited re-takes, etc.) have generated equivalent results; indeed, such a study appears essential for a proper assessment of the efficacy of mathematics placement exam process at PSU.

Assuming that a future study (whether by virtue of a larger sample size, stricter examination procedures or other means) discovers a statistically significant correlation with placement scores and future grade outcomes, there exist a number of attractive methods for building meaningful predictive models. Most commonly, a *multi-variate regression model* may be used to make cogent predictions about the future success of a student placed into a particular math class with a given ALEKS score profile. This regression model could be a 'full' model in the case where each individual sub-category score (including, naturally the overall score) is used to make a prediction – or, on the contrary, a 'reduced' model might be used in the event a particular subset of scores appears *most* significant. In the present study we attempted from the outset to build a multi-variate regression model based upon combinations of subsets of ALEKS scores, but none of these results were deemed statistically significant. The regression model could on the one hand be generated using a standard least-square approach – or, alternatively, depending on the data set, it might be more computationally efficient to 'learn' the regression model by means of a simple *perceptron*. In addition to using a regression-based predictive model, a future study might consider building a *Bayesian Network Classifier* with prior data, conditioned on particular small interval sub-scores as a means to predict future outcomes for 'unseen' test results. Lastly, with sufficient data, it might be possible to successfully use techniques of *Cluster Analysis* (such as the 'nearest neighbor' or *k-means* algorithms) to effectively build new (and perhaps more appropriate) cut-off ranges for course placement.

*Please see additional statistical results attached to this document.